Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	18	"704"/$.ccls. and (speech near3 recogni$) and (((probability near density) or pdf) same (speech near model$1)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 17:56
S1	1	09/866596	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:10
S2	60	(speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:23
S3	26	"704"/$.ccls. and (speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:23
S4	7	(speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620" and ((probability near density) or pdf)	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:21
S5	64	"704"/$.ccls. and (speech near3 process$3) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:48
S6	80	"704"/$.ccls. and (speech near3 process$3) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:49
S7	3	"704"/$.ccls. and (speech near3 process$3) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and (pdf or (probability near density)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:51
S8	4	"704"/$.ccls. and (speech near3 (process$3 or recogni$)) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and (pdf or (probability near density)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:06
S9	869	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:07

S10	151	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:11
S11	3	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and (auto$regress or ar) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:08
S12	12	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and (auto$regress$ or ar) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:08
S13	23	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and ((probability near density) or pdf) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 17:55

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	1	09/866596	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:10
S2	60	(speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:23
S3	26	"704"/$.ccls. and (speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:23
S4	7	(speech near3 recogni$) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620" and ((probability near density) or pdf)	US-PGPUB; USPAT	OR	OFF	2005/01/06 11:21
S5	64	"704"/$.ccls. and (speech near3 process$3) and (MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$))) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:48
S6	80	"704"/$.ccls. and (speech near3 process$3) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:49
S7	3	"704"/$.ccls. and (speech near3 process$3) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and (pdf or (probability near density)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 12:51
S8	4	"704"/$.ccls. and (speech near3 (process$3 or recogni$)) and (((MA or (moving near averag$)) and (AR or (auto$regress$ or (auto near regress$)))) or arma) and (pdf or (probability near density)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:06
S9	869	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:07
S10	151	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:11

S11	3	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and (auto$regress or ar) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:08
S12	12	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and (auto$regress$ or ar) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 14:08
S13	23	"704"/$.ccls. and (speech near3 recogni$) and (speaker near3 (recognition or verification or identification)) and ((noise near3 (suppress$ or reduc$)) or (speech near3 enhanc$)) and ((probability near density) or pdf) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/06 17:55
S14	18	"704"/$.ccls. and (speech near3 recogni$) and (((probability near density) or pdf) same (speech near model$1)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/07 11:16
S15	0	"704"/$.ccls. and (speech near3 recogni$) and (variance with (signal near quality)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/07 11:17
S16	2	"704"/$.ccls. and (speech) and (variance with (signal near quality)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/07 11:17
S17	3	"704"/$.ccls. and (speech) and (variance with ((signal or speech) near quality)) and @ad<"20000620"	US-PGPUB; USPAT	OR	OFF	2005/01/07 11:17

Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler

J.J.Rajan
P.J.W.Rayner
S.J.Godsill

Indexing terms: Parameter estimation, Autoregressive process, Gibbs sampler

Abstract: A nonstationary time series is one in which the statistics of the process are a function of time; this time dependency makes it impossible to utilise standard analytically defined statistical estimators to parameterise the process. To overcome this difficulty, the time series is considered within a finite time interval and is modelled as a time-varying autoregressive (AR) process. The AR coefficients that characterise this process are functions of time, represented by a family of basis vectors. The corresponding basis coefficients are invariant over the time window and have stationary statistical properties. A method is described for applying a Markov Chain Monte Carlo method known as the Gibbs sampler to the problem of estimating the parameters of such a time-varying autoregressive (TVAR) model, whose time dependent coefficients are modelled by basis functions. The Gibbs sampling scheme is then extended to include a stage which may be used for interpolation. Results on synthetic and real audio signals show that the model is flexible, and that a Gibbs sampling framework is a reasonable scheme for estimating and characterising a time-varying AR process.

1 Introduction

In most system identification and estimation techniques it is necessary to assume that the signal is stationary. This requires that the underlying statistics and the model parameters that characterise the process are not dependent on time. However, this assumption is often incorrect for many physical signals encountered in speech processing, EEG analysis and seismology. No general mathematical framework for dealing with nonstationary signals exists, and in practice, the problem of time dependency is circumvented by assuming that the process is locally stationary over a relatively short time interval, but globally nonstationary. The assumption that the process does not depart 'too far' from stationarity over a finite time window then allows the application of standard stationary time series analysis techniques to the data over each finite time window. This procedure is acceptable when the time variation of the process is slow, and the number of observations in each finite window is sufficient to allow a reasonable estimate of the desired model parameters. However, from a modelling point of view this approach is clearly sub-optimal, in that any change in the parameter values can only occur at window boundaries.

In this paper, a Bayesian framework for modelling nonstationary processes using time-varying autoregressive models is described. This framework is quite general as it allows any form of time variation in the parameters to be represented as basis functions. The parameters of the TVAR process are estimated using an efficient Gibbs sampling scheme. This sampling scheme is extended to include an interpolation stage, which samples the interpolant from the overall nonstationary model of the process.

2 Time varying AR models

The modelling of nonstationary signals by autoregressive (AR) models with time-varying coefficients is a well studied problem [1–5]. The signal values are projected onto a basis of time functions, thus allowing the global signal to be represented by a set of constant parameters (i.e. the basis coefficients). In effect, this is just a transformation of the nonstationary signal into a different space, where it can be viewed as a linear time-invariant process. The goodness of fit achieved using this technique is heavily dependent on the subspace spanned by the chosen time functions. The schemes proposed by Grenier [1], Subba Rao [3] and Liporace [6] all assume that the pole movements are relatively slow-varying, and can be modelled using a fixed basis of time functions. Standard choices for these time functions include Legendre polynomials, Fourier basis functions, discrete prolate spheroidal sequences (DPSS) and B-splines. The framework is completely general, as it allows any family of time functions to be used, taking advantage of any available *a priori* information about the signal such as 'seasonal' effects or discontinuities.

IEE Proceedings online no. 19971305

Paper received 19th November 1996

J.J. Rajan was with the University of Cambridge and is now with the Canon Research Centre, 20 Alan Turing Road, Surrey Research Park, Guildford GU2 5YF, UK

P.J.W. Rayner and S.J. Godsill are with the Cambridge University Engineering Department, Trumpington Street, Cambridge CB2 1PZ, UK

2.1 Analysis of time-varying AR models

A stationary AR process y of order p is of the form

$$y(n) = -c_1 y(n-1) - c_2(n-2) - \ldots - c_p y(n-p) + e(n) \quad (1)$$

such that the current observation at time $t = i$ is dependent on the weighted sum of the previous p observations (i.e. the observations at $t = i - 1$, $t = i - 2$, ..., $t = i - p$) plus some stationary excitation noise process e. The weights c applied to the previous observations are known as AR coefficients; if these coefficients are constant and the poles are inside the unit circle, the resulting time series is stationary since the statistics of the process are not dependent on time.

If the AR coefficients are functions that are allowed to vary with time, the general model is considerably more flexible, as it can be used to represent particular types of process where the statistics of the time series are dependent on time. Let x be a time-varying AR process of order p:

$$x(n) = -f_1(n-1)x(n-1) - f_2(n-2)x(n-2) - \ldots - f_p(n-p)x(n-p) + e(n) \quad (2)$$

where f_i is the function that represents the ith AR time-varying coefficient, and $e(n)$ is a zero-mean, stationary Gaussian white noise process.

If the functions f_i can be represented by a family of m basis vectors $\{\mathbf{u}_j: j = 1, \ldots, m\}$, the framework used in [1, 2] is repeated, and the process can be written as

$$x(n) = -\sum_{j=1}^{m} a_{1j}u_j(n-1)x(n-1) - \ldots - \sum_{j=1}^{m} a_{pj}u_j(n-p)x(n-p) + e(n) \quad (3)$$

where a_{ij} are time-invariant basis coefficients. The vectors $\mathbf{u}_j$ are assumed to span some vector space that contains the information needed to represent the functions f_i to some arbitrary tolerance. Without loss of generality, assume that the process $\{x(n): n = 1, \ldots, N + p\}$ is observed. Expressing eqn. 3 in matrix notation

$$\begin{aligned} \mathbf{x} &= -\mathbf{X}_1\mathbf{U}\mathbf{a}_1 - \ldots - \mathbf{X}_p\mathbf{U}\mathbf{a}_p + \mathbf{e} \\ &= -\mathbf{Y}_1\mathbf{a}_1 - \ldots - \mathbf{Y}_p\mathbf{a}_p + \mathbf{e} \\ &= -\mathbf{Y}\mathbf{a} + \mathbf{e} \end{aligned} \quad (4)$$

where x is a vector of samples $\{x(i): i = p + 1, \ldots, N + p\}$, $\mathbf{X}_i$ is an N diagonal matrix with signal values from $x(i)$ to $x(N + i - 1)$ along the main diagonal, and U is a $N \times m$ matrix whose columns are the vectors $\mathbf{u}_j$.

If it assumed that the error residuals are Gaussian and white, the Jacobian of the transformation between the error residuals and the data (i.e. $\mathbf{e} \rightarrow \mathbf{x}$) is unity, and the likelihood function is of the standard form

$$p(\mathbf{x}|\mathbf{a}_1, \ldots, \mathbf{a}_p, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma_e^2} \times \left(\mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\sum_{i=1}^{p}\mathbf{Y}_i\mathbf{a}_i + \sum_{i,j=1}^{p}\mathbf{a}_i^T\mathbf{Y}_i^T\mathbf{Y}_j\mathbf{a}_j\right)\right] \quad (5)$$

Within a Bayesian framework, prior distributions are assigned to the TVAR parameters $\mathbf{a}_i$ and σ_e^2 giving an expression for the joint distribution $p(\mathbf{x}, \mathbf{a}_1, \ldots, \mathbf{a}_p, \sigma_e^2)$. However, obtaining parameter estimates from this joint distribution is generally non-trivial and analytical closed form solutions cannot be found. For this reason, it is necessary to resort to numerical techniques to perform the parameter estimation.

3 Gibbs sampler

The Gibbs sampler is a Markov chain Monte Carlo (MCMC) scheme which allows random variates to be sampled from a joint probability density function $p(\omega, \theta, \sigma)$. It is one of the simplest and most flexible sampling techniques available for this task. The MCMC sampling schemes originated in statistical physics and generally require significant computational expense. With the increasing availability of computing power they are beginning to be applied to general problems in engineering and signal processing [7–10].

The basic idea behind the Gibbs sampler is that the problem of obtaining samples from a joint PDF can be broken down to drawing successive samples from a set of PDFs of smaller dimensionality. The application to parameter estimation problems is obvious. If there is a PDF of large dimension for which the parameter values are required which maximise that PDF, then the Gibbs sampler (by drawing samples from the PDF) enables the formation of histograms of the parameters, from which reasonable point estimates can be inferred.

To sample variates from the joint PDF $p(\omega, \theta, \sigma)$ using the above scheme:

(i) Assume a random starting position $(\omega^0, \theta^0, \sigma^0)$

(ii) Draw a sample ω^1 from the conditional PDF $p(\omega|\theta^0, \sigma^0)$

(iii) Replace ω^0 by ω^1 and continue the process by drawing θ^1 from $p(\theta|\sigma^0, \omega^1)$. As soon as a random variate is drawn it is immediately substituted back into the conditional PDF

(iv) Continue this process of successive sampling from the conditional PDFs for N iterations of the parameters

As with other MCMC methods, this sampling scheme requires an initial transient period known as 'burn-in', where the sampling scheme converges. This should be discarded. The length of the burn-in period depends on the dimensionality and *a posteriori* correlations between the different parameters [11]. Samples drawn after this stage can be considered as having been drawn from the joint PDF $p(\omega, \theta, \sigma)$.

The Gibbs sampler satisfies the conditions of *detailed balance* [Note 1] and it can be shown that the joint density is an invariant distribution of the Markov Chain. More detailed information on the Gibbs sampler and its convergence properties can be found in [10, 12–15].

4 Prior distributions on the AR coefficients

To form the joint PDF of all the free parameters in the model specified by eqn. 3, prior distributions must be assigned to the actual TVAR coefficients. These prior distributions may be non-informative or may incorporate relevant information regarding the physical process. Each of these imposes different constraints on the flexibility of the model. The first is a non-informative Gaussian prior with a common variance. The second is a Gaussian smoothing prior which imposes the con-

Note 1: This implies that an invariant distribution exists and is identical to the joint PDF.

straint that the kth order differences of the TVAR coefficients are minimised, and has the effect of smoothing the resulting representations of the AR coefficients.

4.1 Gaussian prior

If no prior knowledge regarding the model parameters is available, it is possible to consider the basis coefficients as independent normally distributed variables with unknown variance in the prior

$$p(\mathbf{a}|\sigma^2) = \prod_{i=1}^{p}\prod_{j=1}^{m}(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left[-\frac{a_{ij}^2}{2\sigma^2}\right] \quad (6)$$

where m is the number of basis vectors required to represent each of the p AR time-varying coefficients.

A standard prior for application to scale parameters (such as variances) is the inverse Gamma (IG) density, of which the non-informative Jeffreys' prior [Note 2] is a special case [16, 17]. The inverse Gamma prior is of the form

$$p(\sigma^2|\alpha,\beta) = \frac{(\sigma^2)^{-(\alpha+1)}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left[-\frac{1}{\sigma^2\beta}\right] \quad (7)$$

The parameters α and β can be chosen to make the inverse Gamma prior on σ^2 diffuse and prevent the term from collapsing to zero. (This is similar to imposing a constraint that excludes the trivial solution.) This is often necessary for *hyperparameters* [18] such as σ^2, although for well conditioned terms such as the error residual variance σ_e^2 a non-informative Jeffreys' prior will normally be satisfactory. Hence, an inverse Gamma prior is assigned to σ^2 and a Jeffreys' prior to σ_e^2. If $\mathbf{a} = [\mathbf{a}_1^T, \ldots, \mathbf{a}_p^T]^T$, the joint posterior probability density can be expressed as

$$p(\mathbf{x},\mathbf{a},\sigma_e^2,\sigma^2) = p(\mathbf{x}|\mathbf{a},\sigma_e)p(\sigma_e^2)p(\mathbf{a}|\sigma)p(\sigma^2) \quad (8)$$

In the Appendix (Section 9.1), the appropriate conditional PDFs of the parameters are obtained from this expression for the joint PDF to facilitate application of the Gibbs sampler.

4.2 Gaussian smoothing prior

In certain applications, one may possess some prior knowledge as to the general nature of the nonstationarity in terms of the variation of the statistics relating to the process. This information is usually expressed in fairly vague terms such as 'smoothness' of the time dependent coefficients due to known physical constraints on the underlying process. For example, if it is known that there exists a physical constraint on the variation of the AR coefficients, such that their time dependence is 'smooth' in the sense that the second-order differences are small (i.e. approximately maximally flat to degree 2), then this information can be incorporated by placing constraints on the model parameters a_i. The $(N-2) \times N$ matrix $\mathbf{D}$ is defined such that $\mathbf{Dx} = x(n) - 2x(n-1) + x(n-2)$ is the vector of second order differences.

The smoothness constraint applied to the jth basis vector, in the representation of the ith TVAR coefficient, can be expressed in terms of the probability density function of the basis coefficient, conditioned on some regularisation constant λ_{ij}. Hence

$$p(\mathbf{a}_i|\lambda_{i1},\ldots,\lambda_{im}) = (2\pi)^{-\frac{1}{2}}|\mathbf{R}_i^{-1}|^{\frac{1}{2}}\exp\left[-\frac{1}{2}\mathbf{a}_i^T\mathbf{R}_i^{-1}\mathbf{a}_i\right] \quad (9)$$

where $\mathbf{R}_i^{-1} = 2\boldsymbol{\Sigma}_i^T\mathbf{U}^T\mathbf{D}^T\mathbf{D}\mathbf{U}\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_i$ is a $m \times m$ diagonal matrix with terms $\{\sqrt{\lambda_{ij}}: j = 1, \ldots, m\}$ along the main diagonal.

The regularisation constants λ_{ij} describe the ratio of the goodness of fit to the degree of smoothing the constraint applies to the reconstructed signal [19]. This explicitly measures the required degree of constraint, in terms of the actual smoothness provided by each basis vector included in the reconstructed signal. In certain applications, values for λ_{ij} may be known as functions of the data length. The use of the smoothness constraint ensures that high frequency basis vectors do not have an excessive effect on each of the AR coefficient representations, yet at the same time provides a reasonable approximation to the actual time variations of the AR coefficient.

The physical interpretation of the regularisers is as inverse variance terms which act as a set of independent weights on each of the basis coefficients. A possible prior distribution to apply on the λ_{ij} is the Gamma distribution given by

$$p(\lambda_{ij}|\alpha_0,\beta_0) = \frac{\lambda_{ij}^{\alpha_0-1}}{\beta^{\alpha_0}\Gamma(\alpha_0)}\exp\left[-\frac{\lambda_{ij}}{\beta_0}\right] \quad (10)$$

The α_0 and β_0 parameters are chosen so that the Gamma PDF is diffuse and has a mean of one. Since the regularisers act as independent weighting functions on the basis coefficients, it is reasonable to assign prior PDFs on the regularisers that are centred around unity. This assumes that there is no prior information as to the degree of regularisation required by each basis coefficient.

The joint probability density function of the data and the parameters can be written as

$$p(\mathbf{x},\mathbf{a},\lambda_{11},\ldots,\lambda_{1m},\ldots,\lambda_{pm},\sigma_e) = p(\mathbf{x}|\mathbf{a},\sigma_e)p(\sigma_e^2) \times \prod_{i=1}^{p}p(\mathbf{a}_i|\lambda_{i1},\ldots,\lambda_{im})\prod_{i=1}^{p}\prod_{j=1}^{m}p(\lambda_{ij}|\alpha_0,\beta_0) \quad (11)$$

where the diffuse Gamma prior on the regularisers λ_{ij} prevents the terms from going to zero. In a similar manner to that shown in the Appendix (Section 9.1), the appropriate conditional PDFs of the parameters can be obtained from the joint PDF using the Gibbs sampler.

5 Interpolation of missing data

The problem of interpolating missing samples of a time-varying autoregressive process is considered, with the data samples $\{x(i): 1 \le i \le m\}$ and $\{x(i): m + r + 1 \le i \le N\}$ as observed data, and $\{x(i): m + 1 \le i \le m + r\}$ as missing data. The data vector $\mathbf{x}$ is partitioned so that $\mathbf{y}$ represents the observed data and $\mathbf{z}$ is the missing data. The problem of interpolation can then be posed as follows: draw a sample from the conditional PDF $p(\mathbf{z}|\mathbf{y})$ so as to obtain a representative sample of the missing data, conditioned on the observed data.

The application of the Gibbs sampler to this task allows an interpolant for the missing section of data to be obtained, that is typical of the time-varying AR process, and not one that simply best fits the missing

Note 2: The Jeffreys' prior is an inverse Gamma density in the limit (with appropriate scaling factors) where $\alpha = 0$ and $\beta = \infty$. Assigning the Jeffreys' prior to σ^2 results in the distribution $p(\sigma^2) = 1/\sigma^2$. The Jeffreys' prior is known as an improper prior as it is not a normalised PDF.

section. This is desirable in many practical situations such as audio restoration, where it is evident that interpolation based on a maximum likelihood (ML) or expectation maximisation (EM) restoration of missing samples will often result in an interpolant which is atypical of the underlying process [9, 16]. This reasoning corresponds with the statistical physics interpretation of the estimation of the missing samples. Both maximum likelihood estimation and Gibbs sampling draw variates from the Boltzmann distribution $\{\exp[-(1/T)\log p(\mathbf{x}|\theta, \sigma^2)]\}$ but at different temperatures T. Maximum likelihood corresponds to estimation at $T = 0$, while Gibbs sampling draws variates at a temperature of unity, the ambient temperature [10, 12]. Sampling from the predictive density $p(\mathbf{z}|\mathbf{y})$ therefore results in a typical interpolation, since the interpolant and observed data are at the same temperature $T = 1$.

Some expressions that detail the implementation of the method are derived. By rewriting eqn. 4 the error residuals can be expressed in terms of y and z:

$$\mathbf{e} = \mathbf{A}_y\mathbf{y} + \mathbf{A}_z\mathbf{z} \tag{12}$$

where $\mathbf{A}_y$ and $\mathbf{A}_z$ are lower triangular matrices containing the actual TVAR coefficients $\mathbf{f}_i$. The Jacobian of the transformation of variables from the error residuals to the data ($\mathbf{e} \rightarrow \{\mathbf{y}, \mathbf{z}\}$) is unity and the likelihood function is given by

$$p(\mathbf{z}, \mathbf{y}|\sigma_e^2, \mathbf{f}_1, \ldots, \mathbf{f}_p) = (2\pi\sigma_e^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma_e^2} \times (\mathbf{z}^T\mathbf{A}_z^T\mathbf{A}_z\mathbf{z} + 2\mathbf{z}^T\mathbf{A}_z^T\mathbf{A}_y\mathbf{y} + \mathbf{y}^T\mathbf{A}_y^T\mathbf{A}_y\mathbf{y})\right] \tag{13}$$

By representing the TVAR coefficients using a set of basis functions, the joint PDF of the data (both observed and missing) and the associated parameters can be formed as in eqns. 8–11. However, for the purposes of interpolation only the predictive density $p(\mathbf{z}|\mathbf{y})$ is required. The other parameters such as the AR coefficients and the error residual variances are unimportant [12, 13, 16]. Hence, after the Gibbs sampler has converged, MAP (maximum *a posteriori*) estimates for the other parameters should be substituted back into the expression for the joint PDF to give a first order approximate expression for the predictive density

$$p(\mathbf{z}|\mathbf{y}) \approx p(\mathbf{z}|\mathbf{y}, \theta_{MAP}, \sigma^2_{MAP}) \tag{14}$$

This can then be used to obtain an approximate sample for the missing data z. In the Appendix (Section 9.2) the conditional PDF of the missing data z is derived from the joint PDF given in eqn. 13.

6 Results

6.1 Gaussian prior

A synthetic example describing the application of the Gibbs sampler to the problem of estimating TVAR coefficients is presented. Choosing a relatively simple problem for which the correct set of basis functions is known beforehand allows one to verify that the Gibbs Sampler does indeed converge, and draws samples from the correct conditional probability distributions. It should be noted that even for this simple problem, a closed form analytical solution does not exist [1, 2, 20]. The Gibbs sampler is applied to a data set consisting of 2000 samples of a synthetically generated time-varying autoregressive process of order 5. The functional forms (t is sampled uniformly on the interval $[0, 2\pi]$) of the TVAR coefficients are given by $f_1(t) = 0.6\sin(2t)$, $f_2(t) = -0.4\cos(t)$, $f_3(t) = -0.3\cos(2t)$, $f_4(t) = 0.6\sin(t)$, and $f_5(t) = -0.7\cos(t)$ respectively. A Fourier basis set consisting of the five functions $\mathbf{U}_f = [1, \sin(t), \cos(t), \sin(2t), \cos(2t)]^T$ was used. Fig. 1 shows an example of the true TVAR coefficient and its corresponding estimate determined using the Gaussian prior on the coefficients.

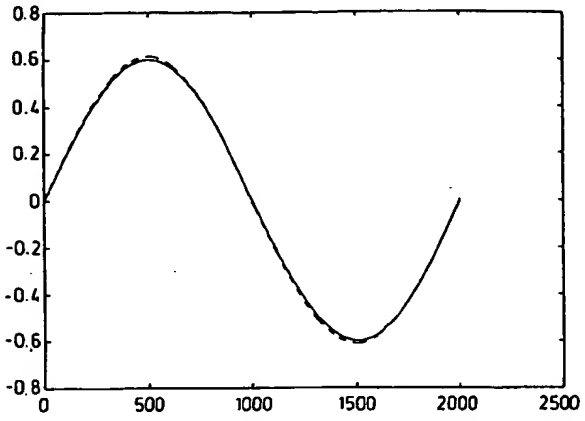

Fig. 1 *4th TVAR coefficient (solid line) and Gibbs sampler estimate (dashed line)*

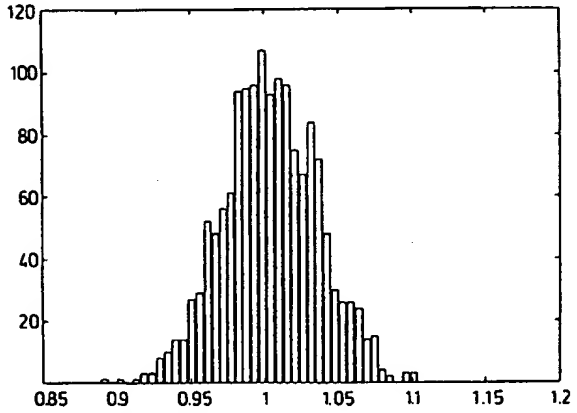

Fig. 2 *Histogram of* σ_e^2
True value = 1.00

Fig. 2 shows the histogram of the error residual variance (the true value is 1.00; the estimated value is 1.004, and the estimated standard deviation is 0.03; the estimates from the histogram are made using a weighted average of the samples in each bin). These results show that the estimates are in good agreement with the true TVAR coefficients $f_i(t)$. In this example, 2500 iterations of the Gibbs sampler were performed. Conservatively, the first 1000 were discarded as 'burn-in'. It should be noted that in practice, the parameter values appeared to converge very quickly, usually within 15 to 20 iterations.

6.2 Gaussian smoothing prior

Applying the Gaussian smoothing prior to the data in Section 6.1 gave almost identical results. This is to be expected, since the smoothness of the TVAR coefficients is primarily dictated by the smoothness of the basis functions used. In this case, the five Fourier basis functions were of low frequency, so the effect of the smoothing constraint was not appreciable.

To demonstrate the Gaussian smoothing prior, the TVAR coefficients were estimated for the same data, but using a different basis set U_h, consisting of the five Fourier basis vectors plus another five (arbitrary) high frequency basis vectors. Figs. 3 and 4 show the representation of the fourth TVAR coefficient using the Gaussian prior and Gaussian smoothing prior respectively. It is clear from these figures that the Gaussian smoothing prior representation is less 'noisy', in that it attempts to suppress the effect of the high frequency basis functions. In practice, this will be of little importance if the set of basis vectors chosen contains only low frequency basis functions.

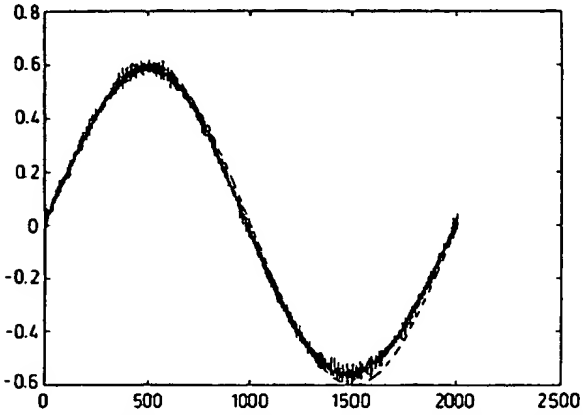

Fig. 3 *4th TVAR coefficient (dashed line) and estimate (solid line) using basis set U_h and Gaussian prior*

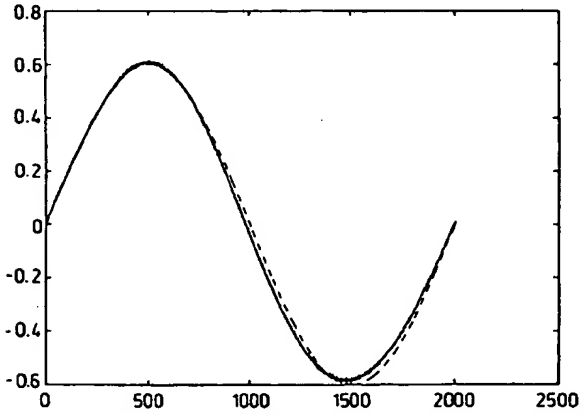

Fig. 4 *4th TVAR coefficient (dashed line) and estimate (solid line) using basis set U_h and Gaussian smoothing prior*

6.3 Interpolation

6.3.1 Synthetic AR(5) data: The problem of interpolation is considered with regard to the time-varying AR(5) process described in Section 6.1: 200 samples between n = 801 and n = 1000 inclusive are removed and considered 'missing'. The Gibbs sampler is then applied to the data, performing a full parameter estimation and an interpolation of the missing data. After 5000 iterations (2000 are discarded as 'burn-in'), MAP estimates of the parameters are made (i.e. the values of the basis coefficients a and the error residual variance σ_e^2). An approximate predictive density $p(z|y)$ is obtained by substituting these parameter values back into the conditional PDF for the interpolant, given by eqn. 28. A sample from the predictive density is chosen as a typical representation of the missing data.

Fig. 5 shows the interpolant obtained using the Gibbs sampler method described above. The estimate of the residual error variance σ_e^2 is 0.9949, and there is no discernible difference between the representations of the TVAR coefficients and the Gibbs sampler estimates obtained without interpolation.

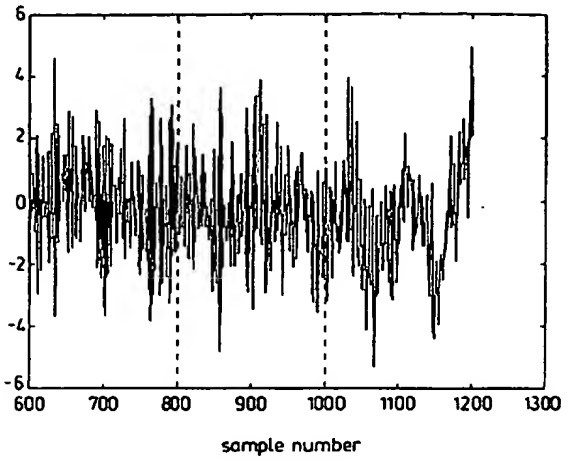

Fig. 5 *Synthetic data: Gibbs sampler restoration*

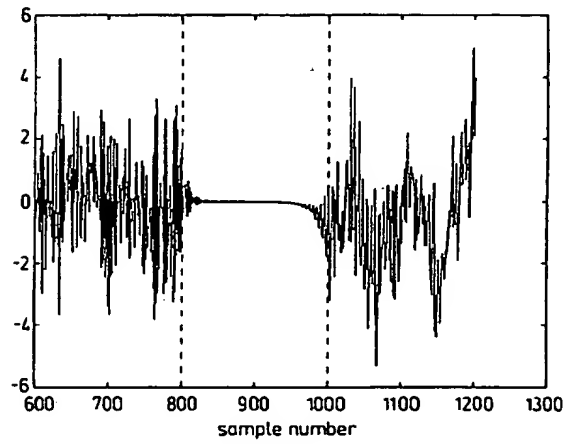

Fig. 6 *Synthetic data: ML restoration*

For comparison, an ML (maximum likelihood) estimate [21, 22] of the interpolant is also determined as shown in Fig. 6. In this case, σ_e^2 is estimated as 0.8818 and the representations of the TVAR coefficients are inferior to those obtained using the Gibbs sampler; a plot of the ML estimate of the 4th TVAR coefficient is given in Fig. 7 (cf. Fig. 1).

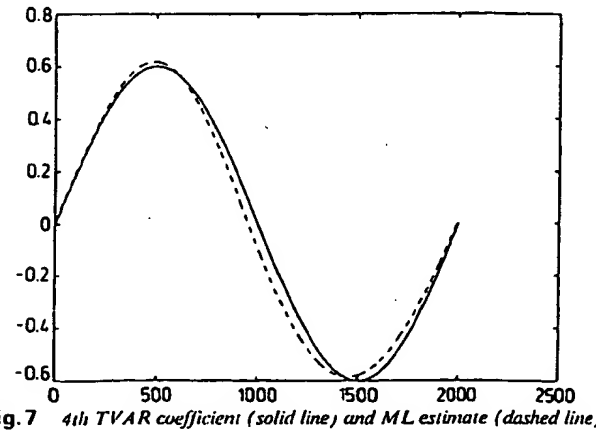

Fig. 7 *4th TVAR coefficient (solid line) and ML estimate (dashed line)*

Figs. 5 and 6 show that the Gibbs sampler interpolant is more typical of the underlying time-varying process than that obtained using the ML method.

6.3.2 Chirp data: The benefits of using the Gibbs sampler for estimating the basis coefficients of a time-varying AR model have been illustrated. However, is there any advantage in using a time-varying AR model, as opposed to a stationary AR model? Consider a 'noisy chirp' signal $x = \sin(\beta t^2) + v$, where the chirp rate $\beta = \pi$, t is sampled uniformly on the interval $[0, 2\pi]$ and v is Gaussian white noise with a standard deviation of 0.1. The signal is clearly nonstationary. One hundred samples are removed, starting with sample $n = 700$, and the Gibbs sampler is applied, assuming a stationary AR model of order 30 in one case and a time-varying AR model (using five Fourier basis vectors to represent the TVAR coefficients) of order 30 in the other. The interpolants formed are shown in Figs. 8 and 9 respectively, and the interpolant in Fig. 9 using the time-varying AR is clearly a better interpolation, owing to the improved modelling of the data.

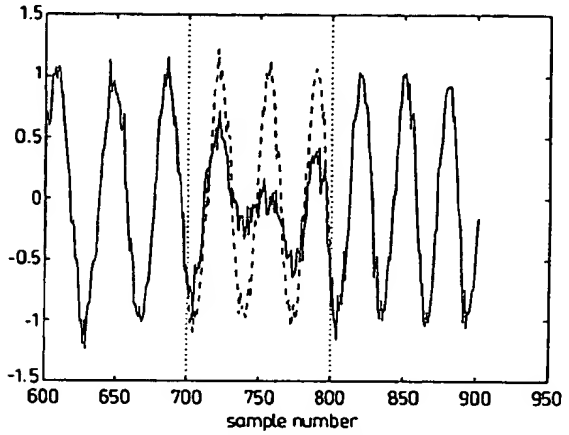

Fig. 8 *Interpolation of chirp using a stationary AR model and Gibbs sampler*
- - - - true chirp signal
——— Gibbs sampler interpolants
Interpolated area is within dotted vertical lines

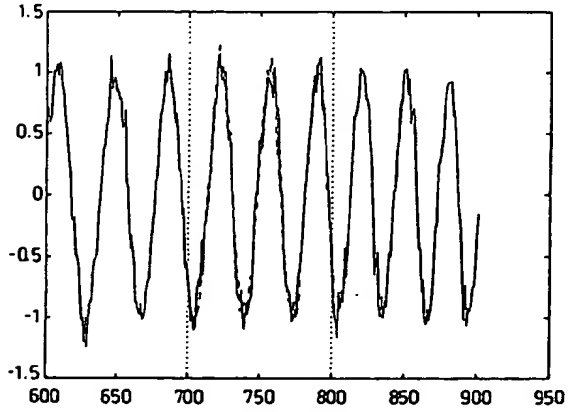

Fig. 9 *Interpolation of chirp using a time-varying AR model and Gibbs sampler*
- - - - true chirp signal
——— Gibbs sampler interpolants
Interpolated area is within dotted vertical lines

6.3.3 Audio data: Audio data are nonstationary but can be considered stationary over a relatively small time interval [16]. In general, the data are modelled using autoregressive models over the time interval during which the data can be considered stationary. As the time variation of the data is relatively slow, the use of a time-varying autoregressive model would be useful in order to model large sections of data and represent the underlying process more accurately. Fig. 10 shows 3000 samples of real audio data (solo vocal music sampled at 44.1 kHz) that are clearly nonstationary.

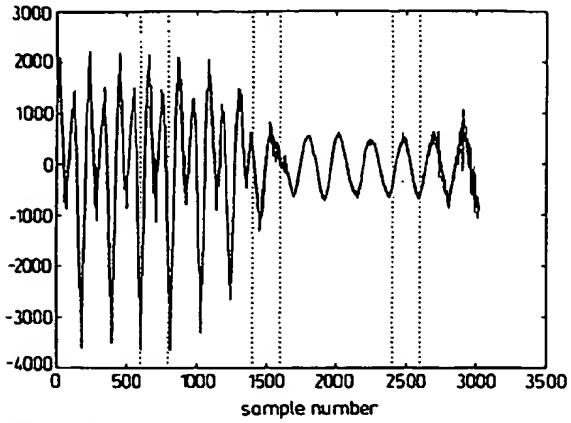

Fig. 10 *Audio data*
Dotted vertical lines indicate interpolation sections

A time-varying autoregressive model of order 25 is applied to the data shown in Fig. 10. A basis set consisting of the first eight Legendre polynomial basis functions is chosen to represent the TVAR coefficients. The choice of basis set is problem-dependent. Initially, a Fourier basis set was chosen but the convergence and variance properties of the basis coefficients were found to be inferior to Legendre polynomial basis functions. The choice of model order was made by plotting a curve of approximate mean squared residual error against model order, and choosing the model order where it appeared that the mean squared error levelled off. This model order selection stage could be formulated within a Bayesian procedure [20, 23].

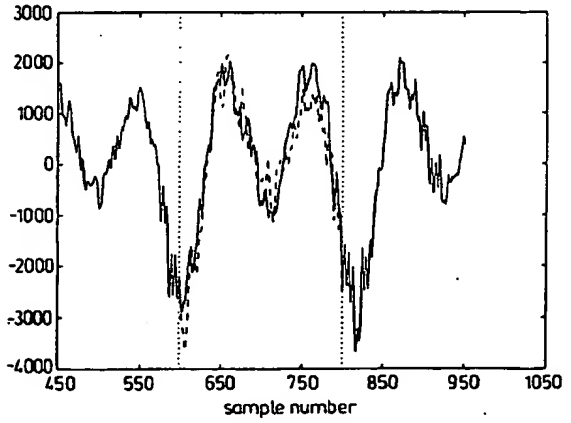

Fig. 11 *Gibbs sampler restoration (solid line) and true data (dashed line)*
$N_0 = 600$; $L = 200$

Three sections, each of L samples, are removed from the data block, as shown by the dotted vertical lines in Fig. 10. Each missing block consists of L samples from $n = N_0 + 1$ to $n = N_0 + L$ inclusive. A Gibbs sampler with a Gaussian smoothing prior is used to estimate the joint density of the three sections of missing data and the unknown model parameters. Figs. 11–13 show the restorations of the $L = 200$ blocks of missing data

at $N_0 = 600$, $N_0 = 1400$ and $N_0 = 2400$ respectively. The restorations appear to 'fit' the missing section in each case, and are typical of the observed data in that local vicinity. This demonstrates that a time-varying AR model can successfully represent an entire nonstationary data block as shown in Fig. 10.

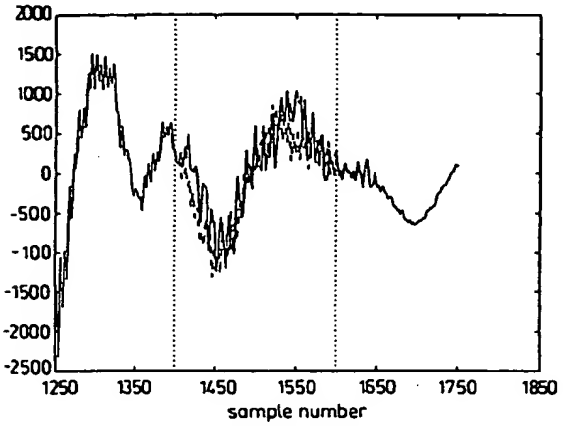

Fig. 12 *Gibbs sampler restoration (solid line) and true data (dashed line)*
$N_0 = 1400$; $L = 200$

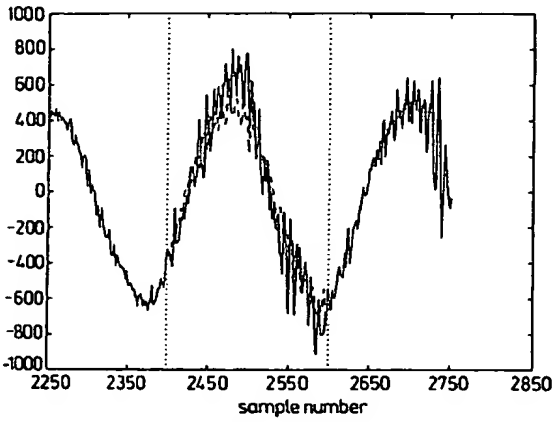

Fig. 13 *Gibbs sampler restoration (solid line) and true data (dashed line)*
$N_0 = 2400$; $L = 200$

7 Conclusion

The problem of applying the Gibbs sampler to the estimation of parameters for time-varying AR processes has been considered. Using a Bayesian approach, the prior on the time-varying autoregressive coefficients can be assigned so as to incorporate constraints such as the degree of smoothness of the parametric variation. Two priors on the time-varying coefficients have been considered: a Gaussian prior and a Gaussian smoothing prior which constrains the coefficient variations to be smooth to the kth degree.

The Gibbs sampler has been applied to several examples of synthetic and real data and the scheme is found to be an effective method for characterising these hierarchical models. In addition, a simple extension to the sampling scheme allows an interpolation stage to be included. The type of nonstationary data resulting from time-varying AR processes, thought to be common in many engineering applications such as speech and audio processing, cannot be modelled or characterised using standard analytical techniques. However, applying the Gibbs sampling scheme to the problem allows histograms of each of the free parameters to be determined, from which reasonable point estimates can be inferred.

8 Acknowledgment

This work was supported by the Human Capital and Mobility Program of the Commission of the European Communities under Contract ERBCHBICT930523.

9 References

1 GRENIER, Y.: 'Time-dependent ARMA modelling of nonstationary signals', *IEEE Trans.*, 1983, ASSP-31, (4), pp. 899–911
2 CHARBONNIER, R., BARLAUD, M., ALENGRIN, G., and MENEZ, J.: 'Results on AR-modelling of nonstationary signals', *Signal Process.*, 1987, 12, (2), pp. 143–151
3 SUBBA RAO, T.: 'The fitting of nonstationary time series models with time-dependent parameters', *J. Roy. Statist. Soc. Ser. B*, 1970, 32, (2), pp. 312–322
4 HALL, M.G., OPPENHEIM, A.V., and WILLSKY, A.S.: 'Time-varying parametric modelling of speech', *Signal Process.*, 1983, 5, pp. 267–285
5 FLAHERTY, M.J.: 'Application of polynomial splines to a time-varying autoregressive model of speech'. Proceedings of IEEE international conference on *Acoustics, speech and signal processing*, 1988, Vol. 1, pp. 2220–2223
6 LIPORACE, L.A.: 'Linear estimation of nonstationary signals', *J. Acoust. Soc. Am.*, 1975, 58, (6), pp. 1288–1295
7 GEMAN, S., and GEMAN, D.: 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Trans. Patt. Anal. Mach. Intell.*, 1984, 6, pp. 721–741
8 GELFAND, A.E., and SMITH, A.F.M.: 'Gibbs sampling for marginal posterior expectations', *Comm. St. A*, 1990, 20, pp. 1747–1766
9 Ó RUANAIDH, J.J.K., and FITZGERALD, W.J.: 'Interpolation of missing samples for audio restoration', *Electr. Lett.*, 1994, 30, (8), pp. 622–623
10 NEAL, R.: 'Probabilistic inference using Markov Chain Monte Carlo methods'. Technical report CRG–TR–93–1, Department of Computer Science, University of Toronto, 1993
11 RIPLEY, B.: 'Stochastic Simulation' (John Wiley, 1987)
12 Ó RUANAIDH, J.J.K., and FITZGERALD, W.J.F.: 'Numerical Bayesian methods applied to signal processing' (Springer-Verlag, 1996)
13 TANNER, M.A.: 'Tools for statistical inference' (Springer Verlag, 1993)
14 TIERNEY, L.: 'Markov chains for exploring posterior distributions', *Ann. Statist.*, 1994, 22, pp. 1701–1762
15 ROBERTS, G.O.: 'Markov chain concepts related to sampling algorithms' *in* SPIEGELHALTER, D.J., GILKS, W.R., and RICHARDSON, S. (Eds.): 'Markov Chain Monte Carlo in practice' (Chapman and Hall, 1996), pp. 45–57
16 GODSILL, S.G.: 'The restoration of degraded audio signals'. PhD thesis, Cambridge University Engineering Department, 1993
17 BERNARDO, J.M., and SMITH, A.F.M.: 'Bayesian theory, vol. 1' (Wiley, 1994)
18 LINDLEY, D.V., and SMITH, A.F.M.: 'Bayes estimates for the linear model', *J. Roy. Statist. Soc. Ser. B*, 1972, 34, pp. 1–41
19 KITAGAWA, G., and GERSCH, W.: 'A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series', *IEEE Trans. Automatic Control*, 1985, 30, (1), pp. 48–56
20 RAJAN, J.J.: 'Time series classification'. PhD, thesis, University of Cambridge, 1994
21 VASEGI, S.: 'Algorithms for the restoration of archived gramophone recordings'. PhD thesis, Cambridge University Engineering Department, 1988
22 VELDHUIS, R.: 'Restoration of lost samples in digital signals' (Prentice-Hall, 1990)
23 RAJAN, J.J., and RAYNER, P.J.W.: 'Bayesian model order selection for the Karhunen-Loève transform and the singular value decomposition'. Proceedings of IEEE international conference on *Acoustics, speech and signal processing*, 1994, Vol. 4, pp. 393–396
24 PRESS, W.H., TEUKOLSKY, S.A., FLANNERY, B.P., and VETTERLING, W.T.: 'Numerical recipes in C – the art of scientific computing' (Cambridge University Press, 1992)

10 Appendix: Derivation of conditional probability density functions

The basic model of the time-varying AR process is assumed

$$x(n) = -f_1(n-1)x(n-1) - f_2(n-2)x(n-2) - \ldots - f_p(n-p)x(n-p) + e(n) \quad (15)$$

where $e(n)$ is stationary Gaussian white noise. Each of the TVAR coefficients $\mathbf{f}_i$ is represented by a set of basis functions $\{\mathbf{u}_j;\ j = 1, ..., m\}$ or by a stochastic process $\mathbf{r}_i$. The residual error sequence e can be written as

$$\mathbf{e} = \mathbf{x} + \mathbf{X}_1\mathbf{f}_1 + ... + \mathbf{X}_p\mathbf{f}_p \tag{16}$$

In all cases, the Jacobian of the transformation between the error residuals and the data (i.e. $\mathbf{e} \rightarrow \mathbf{x}$) is unity and the likelihood function is in the standard Gaussian form

$$p(\mathbf{x}|\mathbf{f}_1, ..., \mathbf{f}_p, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma_e^2} \times \left(\mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\sum_{i=1}^{p}\mathbf{X}_i\mathbf{f}_i + \sum_{i=1}^{p}\sum_{j=1}^{p}\mathbf{f}_i^T\mathbf{X}_i^T\mathbf{X}_j\mathbf{f}_j\right)\right] \tag{17}$$

The parametric form of the conditional PDFs for each of the parameters is of a similar form to the likelihood and prior distributions. This is a property of conjugate priors [17]. The conditional PDFs for the model parameters are Gaussian and those of all the remaining variances are inverse Gamma distributions, for which well known sampling methods exist [24].

10.1 Gaussian prior

The joint PDF of the data and the parameters is given by

$$p(\mathbf{x}, \mathbf{a}, \sigma_e^2, \sigma^2) = p(\mathbf{x}|\mathbf{a}, \sigma_e^2)p(\sigma_e^2)p(\mathbf{a}|\sigma^2)p(\sigma^2|\alpha_0, \beta_0) \tag{18}$$

$$= (2\pi\sigma_e^2)^{-\frac{N}{2}} \times \exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{Y}\mathbf{a} + \mathbf{a}^T\mathbf{Y}^T\mathbf{Y}\mathbf{a})\right]\frac{1}{\sigma_e^2} \times (2\pi\sigma^2)^{-\frac{pm}{2}} \exp\left[-\frac{1}{2\sigma^2}\mathbf{a}^T\mathbf{a}\right] \times \frac{(\sigma^2)^{-(\alpha_0+1)}}{\beta^{\alpha_0}(\alpha_0)} \exp\left[-\frac{1}{\sigma^2\beta_0}\right] \tag{19}$$

10.1.1 Conditional density for the error residual variance σ_e^2:

$$p(\sigma_e^2|\mathbf{x}, \mathbf{a}) \propto (\sigma_e^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma_e^2}E\right]\frac{1}{\sigma_e^2} \tag{20}$$

$$\propto (\sigma_e^2)^{-(\frac{N}{2}+1)} \exp\left[-\frac{1}{\sigma_e^2}\frac{E}{2}\right] \tag{21}$$

where $E = \mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{Y}\mathbf{a} + \mathbf{a}^T\mathbf{Y}^T\mathbf{Y}\mathbf{a}$. The conditional PDF for σ_e^2 is in the form of an inverse Gamma (IG) density with $\alpha = N/2$ and $\beta = 2/E$.

10.1.2 Conditional density for the basis coefficients a:

$$p(\mathbf{a}|\mathbf{x}, \sigma_e^2, \sigma^2) \propto \exp\left[-\frac{1}{2\sigma_e^2}(2\mathbf{x}^T\mathbf{Y}\mathbf{a} + \mathbf{a}^T\mathbf{Y}^T\mathbf{Y}\mathbf{a}^T)\right] \times \exp\left[-\frac{1}{2\sigma^2}\mathbf{a}^T\mathbf{a}\right] \tag{22}$$

$$\propto \exp\left[-\frac{1}{2\sigma_e^2}\left(2\sum_{i=1}^{p}\mathbf{x}^T\mathbf{Y}_i\mathbf{a}_i + \sum_{i=1}^{p}\sum_{j=1}^{p}\mathbf{a}_i^T\mathbf{Y}_i^T\mathbf{Y}_j\mathbf{a}_j\right)\right] \times \prod_{i=1}^{p}\exp\left[-\frac{1}{2\sigma^2}\mathbf{a}_i^T\mathbf{a}_i\right] \tag{23}$$

Computationally, it is quicker to sample the basis vectors $\mathbf{a}_i$ separately, rather than jointly as a concatenated vector a. However, the basis coefficients $\mathbf{a}_q$ that represent the qth TVAR coefficients are dependent on the other basis coefficients $\mathbf{a}_{-q} = \{\mathbf{a}_i;\ i = 1, ..., p;\ i \neq q\}$. Thus the basis coefficients can be be sampled in blocks that represent each TVAR coefficient.

$$p(\mathbf{a}_q|\mathbf{x}, \mathbf{a}_{-q}, \sigma_e^2, \sigma^2) \propto \exp\left[-\frac{1}{2\sigma_e^2}\left(2\mathbf{x}^T\mathbf{Y}_q\mathbf{a}_q + 2\sum_{\substack{i=1\\(i\neq q)}}^{p}\mathbf{a}_q^T\mathbf{Y}_q^T\mathbf{Y}_i\mathbf{a}_i + \mathbf{a}_q^T\mathbf{Y}_q^T\mathbf{Y}_q\mathbf{a}_q\right) - \frac{1}{2\sigma^2}\mathbf{a}_q^T\mathbf{a}_q\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{2\mathbf{a}_q^T}{\sigma_e^2}\left(\mathbf{Y}_q^T\mathbf{x} + \sum_{\substack{i=1\\(i\neq q)}}^{p}\mathbf{Y}_q^T\mathbf{Y}_i\mathbf{a}_i\right) + \mathbf{a}_q^T\left(\frac{\mathbf{Y}_q^T\mathbf{Y}_q}{\sigma_e^2} + \frac{\mathbf{I}}{\sigma^2}\right)\mathbf{a}_q\right)\right] \tag{24}$$

The conditional density of the basis coefficients $\mathbf{a}_q$ is a multivariate Gaussian. Sample from a Gaussian distribution with mean $\mu = -\mathbf{A}^{-1}\mathbf{b}$ and covariance matrix $\mathbf{C} = \mathbf{A}^{-1}$ where

$$\mathbf{A} = \left(\frac{\mathbf{Y}_q^T\mathbf{Y}_q}{\sigma_e^2} + \frac{\mathbf{I}}{\sigma^2}\right) \tag{25}$$

$$\mathbf{b} = \frac{1}{\sigma_e^2}\left(\mathbf{Y}_q^T\mathbf{x} + \sum_{\substack{i=1\\(i\neq q)}}^{p}\mathbf{Y}_q^T\mathbf{Y}_i\mathbf{a}_i\right) \tag{26}$$

10.1.3 Conditional density for the hyperparameter σ^2:

$$p(\sigma^2|\alpha_0, \beta_0) \propto (\sigma^2)^{-\frac{pm}{2}} \exp\left[-\frac{\mathbf{a}^T\mathbf{a}}{2\sigma^2}\right](\sigma^2)^{-(\alpha_0+1)} \exp\left[-\frac{1}{\sigma^2\beta_0}\right]$$

$$\propto (\sigma^2)^{-(\frac{pm}{2}+\alpha_0+1)} \exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{a}^T\mathbf{a}}{2} + \frac{1}{\beta_0}\right)\right] \tag{27}$$

This conditional density is in the form of an inverse Gamma density. The hyperparameter σ^2 should be sampled from IG($pm/2 + \alpha_0$, $2\beta_0/2 + (\beta_0\mathbf{a}^T\mathbf{a})$).

10.2 Interpolation of missing data

Without assuming any specific prior on the TVAR coefficients, the joint PDF for the observed and missing data conditioned on the parameters of TVAR coefficient model is given by

$$p(\mathbf{z}|\mathbf{y}, \sigma_e^2, \mathbf{f}_1, ..., \mathbf{f}_p) \propto \exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{z}^T\mathbf{A}_z^T\mathbf{A}_z\mathbf{z} + 2\mathbf{z}^T\mathbf{A}_z^T\mathbf{A}_y\mathbf{y})\right] \tag{28}$$

This conditional PDF is multivariate Gaussian with mean μ and covariance matrix C given by

$$\mu = -(\mathbf{A}_z^T\mathbf{A}_z)^{-1}(\mathbf{A}_z^T\mathbf{A}_y\mathbf{y}) \tag{29}$$

$$\mathbf{C} = (\mathbf{A}_z^T\mathbf{A}_z)^{-1}\sigma_e^2 \tag{30}$$